Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 1772 – 1779

Procedia Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Visualization System for Shopping Path

Yuta Kaneko[a,*], Shinya Miyazaki[b], Katsutoshi Yada[a]

[a] *Data Science Laboratory, Kansai University, 3-3-35 Yamate, Suita, Osaka 564-8680, Japan*
[b] *Department of Architecture, Fukuoka University, 8-19-1 Nanakuma, Jonan, Fukuoka 814-0180, Japan*

Abstract

The purpose of this research paper is to render the customer-shopping path and customer existence probability visible such that people in the marketing field can easily grasp customer behaviors in store. To achieve this, we introduced the customer existence probability, which provides a visual of how long customers stay in each sales floor zone. The visualization system was then applied to customer-shopping path data to provide separate and concurrent displays of customer-shopping paths and the customer existence probability. Concurrent display of customer-shopping paths/customer existence probability is suitable for visualization of longer customer stays in-store, where the customer moves a great deal. This display enables easy assessment of in-store customer rounds including length of time. It also incorporates to the maximum the information volume of customer-shopping path data.

Peer-review under responsibility of KES International

Keywords: RFID Technology; Visualization System; Customer-Shopping Path; Kernel Density Estimation Method; Customer Existence Probability

1. Introduction

There has been a great deal of research[1] in the marketing field attempting to clarify consumer behaviors using conventional POS data. However, because these were consumer behavior models based on buying results, these methods were not conducive to understanding the product purchasing process. In recent years, owing to the rapid development of information machinery including sensor networks, it is now possible to assess customer behavior in detail.

Larson *et al.*[2] applied multivariate clustering to customer-shopping path data in pioneering research to identify typical customer in-store shopping paths. In addition, Hui *et al.*[3] applied the TSP algorithm to customer-shopping path data, clarifying customer in-store movements by way of deviation of the shortest path between sections of a store and customer-shopping path. Yada[4] clarified critical paths taken by good customers, applying string analysis of the order in which sections of the store were visited. From a statistical perspective, there has also been research[5] to identify latent groups of customers using the Poisson regression model.

* Corresponding author.
E-mail address: r148032@kansai-u.ac.jp

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2015.08.287

Research on customer-shopping paths is progressing such that numerous models have been presented to describe the characteristics of customer in-store shopping paths. Though these models serve the purpose of clarifying important characteristics of customer-shopping paths, the rules and patterns derived in this manner cannot be deciphered by marketing staff or store managers who are not accustomed to mathematical models or algorithms.

For these reasons, we needed an approach that would provide important indications of customer-shopping paths that business people can understand. In conventional research, in-store layout is divided into different areas, and the approach was to gather data for each of these different areas. This approach allowed for simple analysis and the application of algorithms accommodating discrete data. However, model precision depends heavily on in-store spacial discretization, and it is clear that the spacial discretization method has not been studied sufficiently.

The objective of this research is to develop a system providing visibility of in-store customer movements so that business people such as marketing personnel can easily assess customer in-store behavior. By applying the kernel density estimation method, the customer existence probability offers visibility of time-spent by customers in-store without dividing up store layout. This method provides important suggestions by clarifying consumer behaviors for the purposes of laymen.

2. Creating a Visualization System

The purpose of this research is to develop a system displaying customer-shopping paths and the customer existence probability based on customer-shopping path data and visualize in-store movement paths and length of time-spent in stores. Though customer-shopping path data is compiled into graph structure data[6], time information is eliminated with shopping path only displays. In order to incorporate time information, we used the kernel density estimation method to render visible time information as customer existence probability.

The following is an overview of the kernel density estimation method. The kernel density estimation methodology is a non-parametric density estimation method. Because it estimates density directly from data, it does not depend on aggregate units such as store sections, and can estimate continuous density distribution. The kernel density estimation method is used in a variety of different fields since it renders spacial density distribution visible. It is an especially popular method for rendering spacial data visible. Setting the total coordinate points of customer positions distributed planimetrically at n, band width at h, and using two-dimensional normal distribution for the kernel density estimation method, we can formulate the existence probability $\hat{f}_h(x, y)$ on planimetric coordinates (x, y)[7].

$$\hat{f}_h(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2\pi h^2} \exp\left\{ -\frac{(x - x_i)^2 + (y - y_i)^2}{2h^2} \right\} \tag{1}$$

Using the customer existence probability, we rendered visible the amount of time customers spent at different sections in the supermarket.

Our visualization system is comprised of two parts: visualizations of customer-shopping path and customer existence probability. This system is designed to display customer-shopping path and customer existence probability concurrently. By assessing customer behavior by customer-shopping path and customer existence probability at the same time, we can gain a new perspective on customer in-store behavior, particularly how customers make their rounds in the store.

3. Application of visualization system to customer shopping-path data

The visualization system devised here makes individual customer-shopping paths and customer existence probability visible. In addition, we discuss customer-shopping path data in combination with POS data, and in light of correlation with products purchased by customers, in-store shopping paths, and customer existence probability.

3.1. Data Overview

This research utilizes customer-shopping path data acquired using RFID technology in a Kanto region supermarket in October 2012. The customer-shopping path data (position information) was collected each second with an RFID tag placed on the shopping cart. In addition, buying history was linked to the cart ID to indicate which products customers bought. A basic overview of this data is shown in Table 1.

Table 1. Data Overview

Stores surveyed	Kanto area supermarkets
Survey period	2012/10/1-10/31
No. of customers	22,262 people
Average sales amount	2,614 yen
Average number of items sold	15 items
Average time in store	19 minutes

The following is an overview of the store layout. Figure 1 is a map of the store layout. The store is divided into 29 fields including specified product categories. There are two entrances/exits to the store, with produce, seafood, prepared foods, meats, and daily deliveries sections in the outer areas. Inside the store are general food items, confectionaries, alcohol, and more. The color-less areas in the following charts are where product shelves, cash registers, etc., are placed. Customers do not enter these areas.

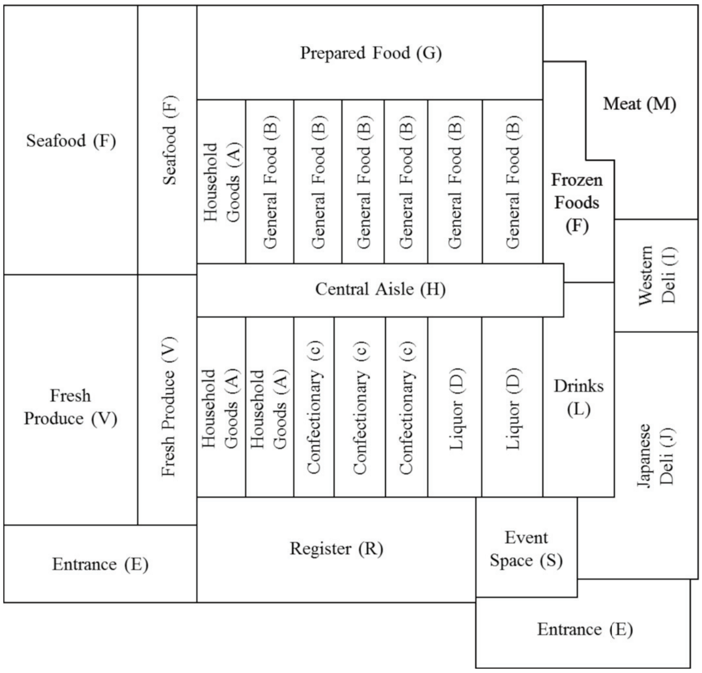


Fig. 1. Grocery Store Layout

3.2. *Visualization of Customer-Shopping Path*

Figure 2 shows customer-shopping path for two customers shopping for a short time. (a) is a shopping path where the customer is walking outside the store, while (b) represents a shopping path inside the store. Customer (a) enters the store and walks past the fresh produce, seafood, prepared foods, meats, and daily deliveries in a clockwise direction. Customer (b) enters the store and goes first to produce, then moves to general foods, meat, and confectioneries. Both customers (a) and (b) were in the store for approximately ten minutes before checking out, which is about half the average length of time a customer spends in this store shopping (about 19 minutes). Therefore both customers (a) and (b) were in the store only for a short period. As shown in the figure, customer (a) did not return to any sections visited

Fig. 2. Visualization of customer-shopping path, (a) customer walks around external area of store; (b) customer walks around inside of store.

previously, while (b) went back to the general foods section just once. Customers who stay in the store for only a short period rarely go back to the same section; instead they tend to quickly pick up the product they want and check out.

We now bring our attention to products purchased by the customer. Customer (a) purchased 43 items for a total of 7,685 yen while customer (b) purchased 11 products for a total of 2,150 yen. Comparing the two, customer (a) purchased roughly four times more items than customer (b), and spent more than three times more money. Customers who shop in the outer part of the store pass by sections where the store makes significant sales, such as produce, seafood, and prepared foods–which means more money is spent. As shown here, the customer-shopping path has a close correlation with sales. Therefore, the ability to visualize the customer-shopping path is extremely important in designing store layout with a view to generating higher sales.

Data appropriate to the visualization of the shopping path is brief-shopping session data characterized by minimal duplicated movement. Where customers move back to the same section numerous times, resulting in multiple overlapping shopping paths, the identity is lost. In addition, visualization of the shopping path clarifies customer rounds made inside the store as well as movements between sections, but does not allow for visualization of the length of time spent in a section. The following sections show visualization of time spent in each section based on customer-shopping path data utilizing the kernel density estimation method.

3.3. *Visualization of the Customer Existence Probability*

In Chapter 2, we formulated the in-store customer existence probability using the kernel density estimation method. Figure 3 shows the visualization of the customer existence probability as noted in definition (1). In the figure, high probability density is shown in red, while blue shows low probability density. (a) shows a density concentrated in the confectionary section, indicating that customers stayed for longer periods in the confectionaries section. In (b), density was distributed, meaning that customers stayed for a time in individual sections as they made the rounds of the different sections. Customer (a) stated in the store for 70 minutes, while customer (b) stayed for 73 minutes, which meant that nearly three times the overall average of 19 minutes was spent on shopping.

Turning our attention to the items purchased, customer (a) bought a total of 20 items including confectionaries and vegetables for a total of 3,926 yen, while customer (b) purchased six items including water and bananas for a total of 1,453 yen. Customer (a) purchased products from high-density sections, which can be said to be related to the customer existence probability and "basket" products. However, the number of products purchased by customer (b) at just six was smaller than the total purchased by (a); though these products were sold in high-density sections,

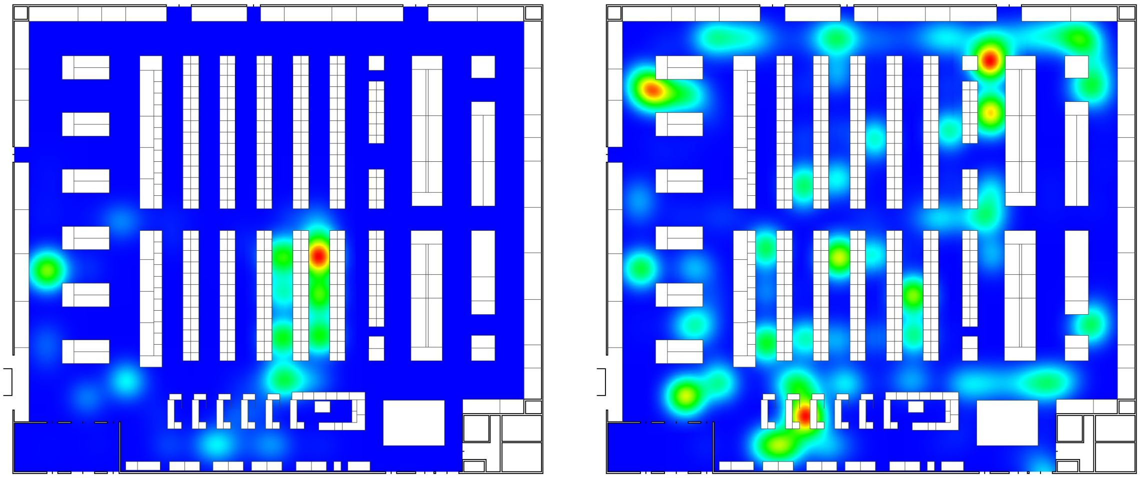


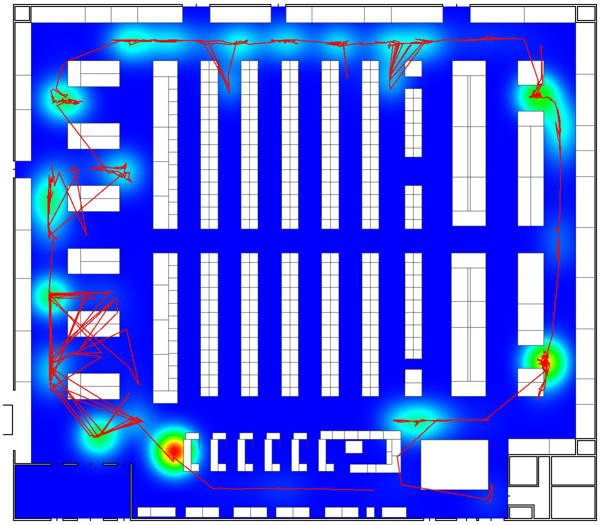
Fig. 3. Visualization of customer existence probability, (a) density distribution concentrated in one area of store; (b) density distributed in different areas of store.

we cannot state that there is a correlation between customer existence probability and "basket" items. Based on this, customer (b)'s purchases represent a pattern of consumer behavior characterized by the customer's rounds in the store, rather than purchasing patterns per se.

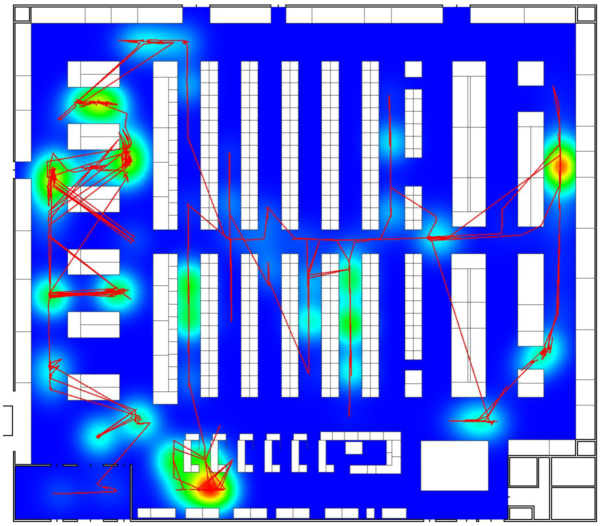
The customer existence probability accurately reflects information on time-spent in store, which is impossible with customer-shopping paths–which means we can assess trends at the various sections where customers stop for longer periods. In addition, because the probability density is continuously distributed, it does not depend on aggregate units such as sections, and it can accurately display data information even at the outer edges of the section. Customer-shopping path displays are not very appropriate for customers exhibiting multiple shopping paths; the customer existence probability, however, even with overlapping shopping paths, is convenient because it is capable of displaying superimposed information. Particularly for customer data where customers stop for longer periods, the customer existence probability is appropriate because it indicates length of time as density. In this way, the customer existence probability accurately displays information on length of time at a section, yet this method alone does not incorporate context of movement between sections or information on chronological order. The following is a discussion on the concurrent display of shopping paths and customer existence probability.

3.4. Visualization of customer-shopping path and customer existence probability

Figure 4 is a concurrent display of customer-shopping paths and customer existence probability. As in Figure 2, (a) indicates a pattern where the customer makes the rounds of the outside area of the store, while (b) shows a pattern where the customer does the rounds on the inside of the store. Comparing to Figure 2, Figure 4 indicates greater customer movement in the section, and there are points where customer-shopping paths overlap. Customer (a) enters form the entrance, stopping at the external sections in counterclockwise fashion at the event space, Japanese daily deliveries, meat, prepared foods, seafood, and produce. Customer (b), meanwhile, stops at both the outer and inner sections of the store: produce, seafood, confectionaries, daily Japanese deliveries, meats, and daily goods. Customer (a) stayed in the store for 37 minutes, while customer (b) stayed for 38 minutes–roughly the same amount of time. Both stayed nearly double the average. Analyzing by the individual sections that the customer stopped at, customer (a) stopped at numerous sections including daily deliveries, meats, and fish, while customer (b), stopped many times at the produce, fish, and meat sections. Both customers (a) and (b) moved numerous times between sections in the outer part of the store, raising the customer existence probability. Customers tend to stop for longer and move between sections in the outer area of the store, where sales are higher.



(a) Customer walks around outside areas of store



(b) Customer walks around inside of store

Fig. 4. Visualization of customer-shopping path and customer existence probability, (a) customer walks around outside areas of store; (b) customer walks around inside of store.

Examining products purchased, customer (a) bought a total of 25 items for a total of 4,864 yen while customer (b) purchased six products for a total of 844 yen. Comparing the two, though they were both in the store for approximately the same length of time, customer (a) purchased more than four times the number of items produced by customer (b), and spent nearly six times the money. We believe this is due to the fact that customer (a) visited primarily the outer sections of the store, where sales are high.

As described here, displaying the customer-shopping path and customer existence probability concurrently allows the user to assess customer patterns of making the rounds in a store, including length of time at a section. For example, we can infer that because there is significant customer movement at the produce and fish sections, the customer existence probability is also high. Customer-shopping path data includes customer shopping path and information regarding length of time at a section, but because only one variable–either the customer-shopping path or customer existence probability–can be rendered visible, not all information can be incorporated. This means inferior precision. By rendering both visible, we maximize information precision of customer-shopping path data, and enhance visual understanding of consumer behavior.

3.5. Summary of the Visualization System

Table 2 is a summary of the details of customer-shopping path visibility discussed in this paper.

Table 2. Summary of visualizing shopping path

Visualization method	Advantages	Data pertaining to visibility
Shopping Path	· User can assess context of customer movement	· Brief-shopping session data
Customer Existence Probability	· Allows for assessment of where customers tend to stop in stores	· Long-shopping session data where movements are duplicated
	· Not dependent on individual store aggregate units; allows for accurate information in all areas of store	· Data characterized by customers staying in store longer/less movement
Shopping Path + Customer Existence Probability	· Allows for assessment of customer movements in stores	· Data characterized by customers staying in store for longer periods/more movement

First, customer-shopping paths are applicable to visualizing brief-shopping session data characterized by minimal duplicated movement. Because the data is displayed linearly, it is superior for helping the user to assess chronological order of movement around the section. However, it cannot visualize length of stay in store. Though the customer existence probability is ideal on the other hand for data characterized by a longer presence in the store where there is significant duplication of data; or where customer stay in store is longer and there is little movement. Though it can visualize length of time in store, it is unable to handle information on movement around the section. In conclusion, concurrent display of customer-shopping path and customer existence probability is superior for data characterized by long periods in store and where there is greater movement. Because it can visualize customer-shopping paths and length of time in store, it can reflect the maximum information volume that is characteristic of customer-shopping path data.

4. Discussion

As part of this paper, we developed a system to help users visualizing the customer-shopping path and customer existence probability based on customer-shopping path data. By managing customer in-store movement, we attempt to provide important suggestions for store management. Though conventional systems displayed customer shopping-paths, there were problems visualizing time-spent in store. By rendering the customer existence probability visible with the kernel density estimation method, the user can easily assess time-spent in-store at each section. The visualization system we present here provides an important perspective in helping marketing staff and store managers–individuals not accustomed to mathematical models–to understand customer in-store behavior.

There remain, however, numerous issues related to this research. Because our research focuses on improving visibility of in-store behavior for each customer, we are unable to assess multiple customer group trends. In the future, we need to conduct experiments on how to render multiple customer-shopping paths more visible. In addition, because customer-shopping paths can become very large, an efficient processing system capable of handling large data is required. Also, even customer behavior in the same store can be affected by changes in time of day and season, which means we also need a system that can achieve visibility of changes in chronological order.

Acknowledgements

This work was supported by "Strategic Project to Support the Formation of Research Bases at Private Universities": Matching Fund Subsidy form MEXT (Ministry of Education, Culture, Sports, Science and Technology), 2014-2018.

References

1. P. M. Guadagni and J. D. Little, "A logit model of brand choice calibrated on scanner data", *Marketing Science 2(3)*, pp.203-238(1983).
2. J. S. Larson, E. T. Bradlow and P. S. Fader, "An exploratory look at supermarket shopping paths", *International Journal of research in Marketing 22(4)*, pp.395-414(2005).
3. S. K. Hui, P. S. Fader and E. T. Bradlow, "Research note-The traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality", *Marketing Science 28(3)*, pp.566-572(2009).
4. K. Yada, "String analysis technique for shopping path in a supermarket", *Journal of Intelligent Information Systems 36(3)*, pp.385-402(2011).
5. K. Takai and K. Yada, "Finding latent groups of customers via the Poisson mixture regression model", *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pp.3603-3608(2011).
6. S. K. Hui, P. S. Fader and E. T. Bradlow, "Path data in marketing: An integrative framework and prospectus for model building", *Marketing Science 28(2)*, pp.320-335(2009).
7. S. Miyazaki, T. Washio and K. Yada, "Analysis of Residence Time in Shopping using RFID Data–An Application of the Kernel Density Estimation to RFID", *2011 11th IEEE International Conference on Data Mining Workshops*, pp.1170-1176(2011).